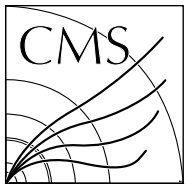
Available on CMS information server

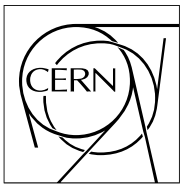
**CMS AN-2008/015**




**The Compact Muon Solenoid Experiment**

# Analysis Note

The content of this note is intended for CMS internal use and distribution only




**1 May 2008**

# Expectations for $t\bar{t} \rightarrow$ dileptons in the early phase of CMS

L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, O. Gutsche

*Fermilab*

F. Golf, J. Muelmenstaedt, M. Norman, A. Yagil, F. Würthwein

*University of California, San Diego*

C. Campagnari, P. Kalavase, V. Krutelyov, D. Kovalskyi, J. Ribnik

*University of California, Santa Barbara*

**Abstract**

We present expectations for the observation of $t\bar{t} \rightarrow$ dileptons ($e$ or $\mu$) in the early phase of CMS running. We define a possible event selection, and we survey the Standard Model sources of dileptons. We find that a clear $t\bar{t}$ signals stands out when selecting two high $P_T$ leptons, missing $E_T$, and $\geq 2$ jets.

# 1 What Is Missing From This Note

- Would be nice to have a table of expected event yields corresponding to Figure 11.
- Plots of kinematical distributions are missing. Need to choose an interesting set and put them in.
- Section 12 (data driven determination of $W+$ jets background) is incomplete. It will remain incomplete until the work is finished. It is not for approval anyway.
- Some of the references need fixing.

# 2 Introduction

The dilepton $t\bar{t}$ final state, where both $W$s from $t \to Wb$ and $\bar{t} \to W\bar{b}$ decay leptonically into $e$ or $\mu$, is the cleanest $t\bar{t}$ final state. In this note we outline a possible early analysis which aims to establish $t\bar{t} \to$ dileptons at CMS and to provide a first measurement of the $t\bar{t}$ cross-section.

The strategy for the analysis described in this note is that of a simple counting experiment: we define an event selection, we count the number of candidates, we compare this number to the number of expected candidates from all non-$t\bar{t}$ Standard Model (SM) sources, and we ascribe the excess of events to $t\bar{t}$. In such a counting experiment it is very important to be able to validate the estimation of the non-$t\bar{t}$ SM contributions. To do this we use control regions where the non-$t\bar{t}$ background is expected to be large and the $t\bar{t}$ signal is small. In practice this can be done most easily by dividing the event sample in jet multiplicity bins. This is because $t\bar{t}$ events tend to have jets from the $b$-quarks in top decay, while most other backgrounds tend to have no jets in the final state. The analysis will then consist of selecting events with two high $P_T$ opposite sign isolated leptons, $\not\!\!E_T$, and some number of jets ($N_{jets}$). We will then test the background prediction in the $N_{jets} = 0$ and 1 bins, and extract the $t\bar{t}$ signal in the $N_{jets} \geq 2$ bins.

In a second phase of this analysis, which is not described here, we will also compare the kinematical distributions of the events with expectations from $t\bar{t}$ and SM backgrounds. By its nature this then becomes a broad model-independent search for New Physics in dileptons.

Note that $b$-tagging is not used in this analysis. This is because the analysis is aimed at the very early data, before the establishment of $b$-tags. Others in the top group are pursuing a similar analysis that exploits $b$-tags[1].

The scope of this note is the following:

- Survey all sources of SM dileptons.
- Identify tools, variables, and techniques that are useful in separating signal from background.
- Define a reasonable set of base requirements for this analysis. These will of course be re-evaluated with real data.
- Understand how to estimate the backgrounds in a data-driven way, whenever possible.
- Survey the various systematic uncertainties for the cross-section measurement, and provide a first estimate of their size.

There are three dilepton final states that we consider: $ee$, $\mu\mu$, and $e\mu$. The $ee$ and $\mu\mu$ final states are more difficult than the $e\mu$ final state because of the Drell-Yan (DY) background, $pp \to \gamma^*/Z^{(*)} \to e^+e^-$ or $\mu^+\mu^-$. The ability to reject the large DY background depends crucially on $\not\!\!E_T$. The $\not\!\!E_T$performance of the detector at startup is very uncertain. Here we will carry out our analysis assuming that the current CMS Monte Carlo gives a good representation of the $\not\!\!E_T$. We are very aware that when real data arrives we may have to redesign the analysis in the $ee$ and $\mu\mu$ final states, and may be even be forced to limit ourselves to $e\mu$.

# 3 Data Sets

We use Monte Carlo data sets from CSA07[2].

## 3.1 Data Set Summary

The $t\bar{t}$, $W$+jets, and $Z$+jets (DY) data sets come from the so-called chowder soup. These events were genereated with Alpgen. They have been reconstructed with three different miscalibration and misalignment scenarios (startup, 10 pb$^{-1}$, and 100 pb$^{-1}$). These data sets have been filtered through the HLT, and have subsequently been skimmed using the two $t\bar{t}$ dilepton skims, `topDiLepton2Electron` and `topDiLeptonMuonX`[3]. Note, however, that miscalibrations were not taken into account at the trigger level (L1 or HLT).

The `topDiLepton2Electron` skimmed data set is formed from the `HLT1ElectronRelaxed` and `HLT2ElectronRelaxed` HLT streams; the `topDiLeptonMuonX` skimmed data set is formed from the `HLT1MuonNonIso`, `HLT2MuonNonIso`, and `HLTXElectronMuonRelaxed` HLT streams. The skims simply require that there be two leptons ($e$ or $\mu$) of $P_T > 20$ GeV in the event. Thus, one expects $ee$ events to come from the `topDiLepton2Electron` skim and $\mu\mu$ events to come from the `topDiLeptonMuonX` skim. On the other hand $e\mu$ events could come from both skims, since they could fire the single electron trigger, in which case they would show up in the `topDiLepton2Electron` skim, or they could fire the single muon trigger or the $e\mu$ trigger, in which case they would show up in the `topDiLeptonMuonX` stream. In our analysis we take care to not double-count $e\mu$ events that appear in both skims.

Some backgrounds, eg, QCD Alpgen backgrounds, were processed in the same way as the $t\bar{t}$, $W$+jets, and $Z$+jets samples, and were then put together in separate soups (stew and gumbo).

Other backgrounds to our analysis, *e.g.*, $WW$, $WZ$, $ZZ$, were generated and processed in CSA07 as so-called *signal samples*. This means that the results of the HLT filters were flagged but not applied, and the events were not skimmed. When processing these samples, we make sure to apply the appropriate trigger requirements. We do not pass these events through the top dilepton skim filter, since its requirements are looser than our selection requirements. These samples were reconstructed with 100 pb$^{-1}$ calibrations.

## 3.2 Cross-sections

We normalize all of our data sets to the cross-sections from CSA07. For the $W$ and Drell Yan in the soups, we use as normalization the information provided by the `CSA07EventWeightProducer` module[4]; these are the cross-sections from Alpgen. For the $t\bar{t}$ sample, the `CSA07EventWeightProducer` module gives the LO cross-section. We renormalize it by the appropriate K-factor. For the diboson samples, we use the information from the Pythia cfg files (NLO cross-sections). The assumed cross-sections as well as the equivalent luminosities for the most important processes are summarized in Table 3.2. All samples are simulated with statistics sufficient to neglect the uncertainty from MC statistics for the assumed data sample of 10 pb$^{-1}$.

| Sample | Cross Section | Equivalent luminosity* |
|---|---|---|
| $t\bar{t}$ | 447×1.87=837 pb [5] | 2.4 fb$^{-1}$ |
| $W$ | 58.2×1.12 = 65.2 nb [5] | 0.2 fb$^{-1}$ |
| Drell-Yan | 5.77×1.12 = 6.46 nb [5] | 0.7 fb$^{-1}$ |
| $WW$ | 114.3 pb | 6.5 fb$^{-1}$ |
| $WZ$ | 49.9 pb | 7.3 fb$^{-1}$ |
| $ZZ$ | 16.1 pb | 8.5 fb$^{-1}$ |

Table 1: Assumed cross-sections and equivalent integrated luminosities for the main Monte Carlo samples used in analysis. In all cases, except $W$ and Drell-Yan, the cross-section is irrespective of decay mode. For the $W$ sample, the cross section is for the sum of $e\nu$, $\mu\nu$, and $\tau\nu$. For the Drell-Yan sample, the cross section is for the sum of $ee$, $\mu\mu$, $\tau\tau$. For samples generated with the Alpgen which are a combination of multiple samples with different parton level requirements the integrated luminosity is given for the subsample with the lowest luminosity.

## 3.3 Limitations of our Monte Carlo samples

Here we list the known limitations of our Monte Carlo samples.

- The statistics of the QCD background samples are limited. It is difficult to use them to estimate the dilepton QCD backgrounds reliably. In addition, it is not clear how reliable these samples are in the first place.
- We are missing the $Wb\bar{b}$ and $Wc\bar{c}$ Monte Carlo samples. These are important samples to model the background where one of the leptons comes from $W$ decay and the other one comes from heavy flavor decays.

Note that the $W+$ jets Monte Carlo sample only inludes gluon and light flavor quarks from the matrix element. Some $b\bar{b}$ and $c\bar{c}$ pairs are then generated in the gluon showers, but the theoretical state of the art estimates of $Wb\bar{b}$ and $Wc\bar{c}$ require the dedicated event samples. Separate studies indicate [1] that contributions from such processes are well within uncertainties on the expected $W+$jets background.

- The Alpgen Drell-Yan sample was generated with dilepton invariant mass above 40 GeV. The skim selection, see Section 3.1, requires both leptons to have $P_T > 20$ GeV, so that naively one would think that the 40 GeV dilepton mass requirement is safe. However, this is not the case for events with jets, where the dilepton pair can be boosted, see Figure 1. Note that since the $Z$ peak is vetoed in our analysis, it is the events with mass away from $M_Z$ that matter to us most.

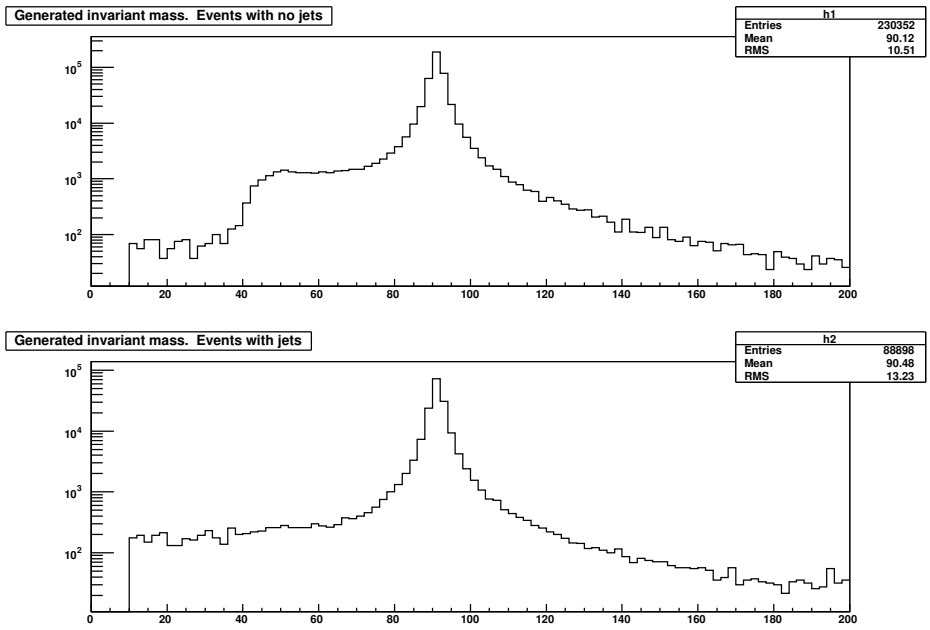



Figure 1: Generated dilepton invariant mass spectrum (GeV) from a pre-CSA07 Pythia Drell Yan $\rightarrow \mu\mu$ sample for events with two leptons with $P_T$ above 20 GeV. The top plot is for event with no reconstructed caloJets of uncalibrated $E_T > 15$ GeV and $|\eta| < 3$. The bottom plot is for events with at least one such caloJet. The Pythia events were generated with dilepton invariant mass above 10 GeV. The normalization of the vertical scale is arbitrary.

# 4 Triggers

Events in our analysis can come on the single lepton triggers or on the dilepton triggers. In principle there is also some acceptance to other triggers, *e.g.*, jet triggers, but these triggers were not in the skims[1)]. The relevant HLT triggers for the CSA07 top dilepton skimmed data sets are listed in Table 4 [6],[3].

| Single $e$ triggers | Single $\mu$ triggers | Dilepton Triggers |
|---|---|---|
| HLT1ElectronRelaxed 17 | HLT1MuonNonIso 16 | HLT2ElectronRelaxed 12,12<br>HLT2MuonNonIso 3,3<br>HLTXElectronMuonRelaxed 10,10 |

Table 2: The single and dilepton triggers in CSA07. Electron relaxed trigger have no explicit isolation requirements at level 1, but all electron triggers have isolation requirements at the HLT.

The efficiency of the various trigger lines on top events have been studied in detail by others in the top group [7]. Here we just quickly look at some of the issues ourselves.

First, we note that the single lepton triggers have thresholds below the skim and analysis threshold of 20 GeV on $P_T$ of the leptons. With one exception noted below, the dilepton triggers are made up of the same basic objects as the single lepton triggers, with the same requirements, but lower thresholds[2)]. Thus, the only advantage of including the dilepton triggers would be if the turn-on curves of the single lepton triggers were shallow enough that addition of the dilepton triggers would significantly increase the trigger efficiency for leptons with $P_T$ near 20 GeV.

The one exception is the $\mu\mu$ trigger, since the single muon trigger is limited to $|\eta| < 2$, while the dimuon trigger extends the coverage to $|\eta| < 2.4$. Thus, the dimuon trigger is the only trigger that can be used to collect events where both muons have $2 < |\eta| < 2.4$.

As argued above, the turn-on curves for the single lepton triggers are of particular interest. These turn-on curves are hard to extract from the $t\bar{t}$ events in the chowder skimmed samples, since the trigger has already been applied. Thus, we work with the $WW$ signal sample, requiring one $W \rightarrow e\nu$ and one $W \rightarrow \mu\nu$ at the generator level. On these events, we measure the efficiency as a function of $P_T$ and $\eta$, see Figure 2. This efficiency is for events where the electron or muon was found offline, passed the muon selection of Section 6.1 or the egamma POG tight selection (see Section 6.2), and was truth-matched to $W \rightarrow e\nu$ or $W \rightarrow \mu\nu$. No isolation requirements were applied.

The muon trigger efficiency curves in Figure 2 look very much like the ones in Reference [7] on $t\bar{t}$ events. However, in the case of electrons, we find a higher efficiency. We note that in the measurements from Reference [7] neither truth match nor electron identification requirements were applied.

From Figure 2 we note that the single lepton triggers have already plateaued at $P_T$ = 20 GeV. Thus we anticipate little additional improvement in efficiency by adding the dilepton triggers.

The overall efficiency for the single lepton triggers are not as high as we would want them to be, particularly in the case of electrons. Work is ongoing in the trigger group to improve the situation.

To measure the impact of the dilepton triggers in our analysis, we consider $t\bar{t}$ events in the chowder soup described in Section 3.1. We select at the generator level events from the dilepton decay chain. We then look at events that pass the base selection of Section 6. We find that addition of the dilepton triggers improves the efficiency by less than 1%. Thus, for simplicity we do not use the dilepton triggers for the analysis described in this note. Our HLT trigger requirements are summarized in Table 4.

# 5 Code Details

We work in CMSSW version 1.6.8. We use the TQAF layer 1 objects using version 168_080131 of the TQAF. We modify the TQAF default configuration as follows:

[1)] We would not have wanted to use them anyway because they would not add very much and would make the analysis more complicated

[2)] This is actually not quite true. The $E/P$ requirement is not applied in the HLT $ee$ trigger for CSA07. In more recent versions of the HLT code, the $E/P$ requirement has been removed from all electron trigger paths, so the statement in the text is correct for those HLT versions.

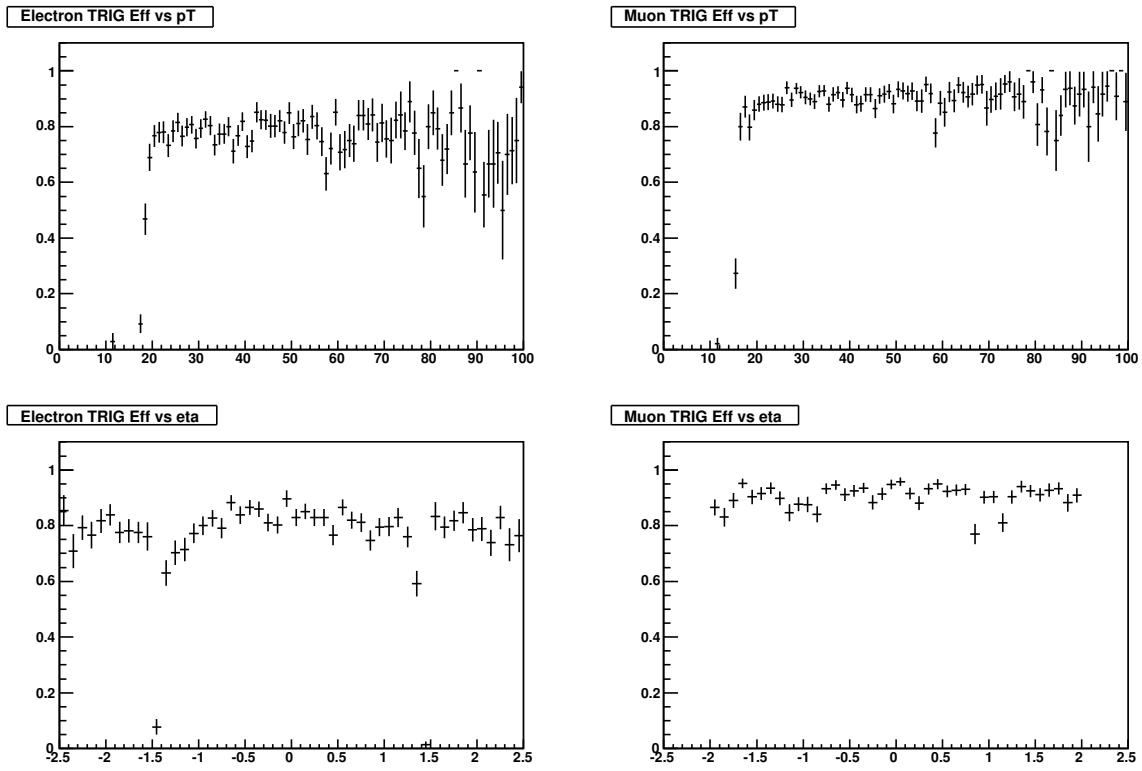



Figure 2: Single electron and single muon trigger efficiency for reconstructed electrons and muons in $WW$ events. See text for details. Muons were required to have $|\eta| < 2$. For the $\eta$ efficiency plots, we require $P_T > 20$.

| $ee$ mode | $\mu\mu$ mode | $e\mu$ mode |
|---|---|---|
| HLT1ElectronRelaxed 17 | HLT1MuonNonIso 16 | HLT1ElectronRelaxed 17<br>OR<br>HLT1MuonNonIso 16 |

Table 3: HLT trigger requirements for this analysis.

- No filters after Layer1XXX producers.
- We do not use the TQAF $\not\!\!E_T$, since we prefer to use $\not\!\!E_T$not corrected for jet energy scale effects, see Section 6.3.
- The TQAF TopJet producer is configured to not do the cleaning of jets from leptons, see Section 6.6.

We do not use the TQAF layer 2, since our analysis is very simple. We just read in the TQAF layer 1 objects, select and count leptons and jets, and look at the $N_{jets}$ and other kinematical distributions.

# 6 Base Selection

Here we briefly describe our event selection. In many cases we examine the effects of misalignments and miscalibrations on our event selection, since this is an analysis that will be carried out with early data. Clearly there is no guarantee that the available misalignments and miscalibration scenarios are realistic. They should be used only as a rough guide. Much of our work in the early stages of data taking will be in commissioning the physics objects.

## 6.1 Muon Selection

We use global muons with $P_T > 20$ GeV. The muon POG group has not yet provided standard selections to be used for Physics. High $P_T$ muons in CMS are quite clean, but there are some backgrounds. We have seen [8] that

about one third of reconstructed global muons in QCD events are not from $b$ or $c$ decays. They are a combination of kaons/pions decays in flight and residual punchthrough. This background can be reduced by approximately a factor of two by applying simple selection requirements.

For the moment our selection requirements come from our own study of single pion, kaon, and muon events [9]. These requirements are:

- Impact parameter from global fit $|d_0| < 2.5$ mm (in the XY plane).
- Number of valid hits from silicon track fit $\geq 7$.
- $\chi^2/dof$ from global fit $< 5$.

The first two requirements are effective against decay in flights. Note that the impact parameter cut is still very loose and could be tightened. A $\Delta Z$ requirement between the track and the vertex could also be useful. The $\chi^2$ requirement reduces the rate of interactive punch-through backgrounds.

The efficiency of these requirements is very high and insensitive to the calibration scenarios, see Figure 3. (Of course, this will have to be evaluated with real data). The efficiency is defined for truth matched muons as the efficiency with respect to the offline reconstruction, *i.e.*, the efficiency for a muon to pass the muon ID requirements given that a reco::muon object was made.

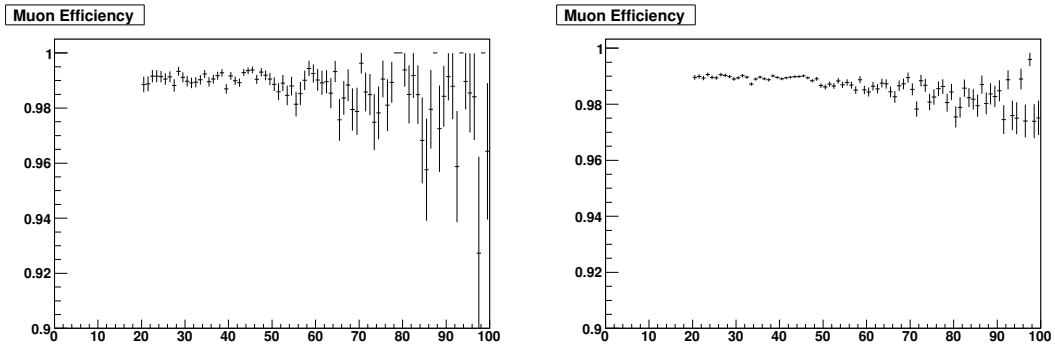



Figure 3: Efficiency for the muon selection described in the text on $Z \rightarrow \mu\mu$ events. Left plot: startup scenario; right plot: 100 pb$^{-1}$ scenario. Note the suppressed zero on the vertical scale. See text for the precise definition of efficiency.

Another handle against QCD backgrounds is isolation. We will discuss it in Section 6.4.

## 6.2 Electron Selection

For electron selection we have examined the three cut-based standard selections from the egamma POG (robust, loose, and tight) [10]. Working with pre-CSA07 data, we found [11] that the tight selection was optimal in terms of efficiency and background rejection. Therefore, we adopt the egamma POG tight selection as our base selection. The reasons for this choice are discussed in Section 7.2.

One concern is the impact of misalignments and miscalibrations on the electron efficiency. In Figure 4 we show the efficiency of the robust, loose, and tight selections as function of electron $P_T$ for different calibration scenarios. The efficiency is defined for truth matched electrons as the efficiency with respect to the offline reconstruction, *i.e.*, the efficiency for an electron to pass the electron ID requirements given that a reco::electron object was made.

From Figure 4 we conclude that the poor calibration scenarios have little effect on the electron efficiency (with respect to the offline reconstruction efficiency). Clearly one cannot claim that the available miscalibration scenarios capture all of the problems that we will face at startup. The robustness of the electron selection efficiency remains an open question which we will have to face with real data.

In previous work we also found that radiative decays *e.g.*, $Z \rightarrow \mu\mu\gamma$, can result in reconstructed electrons when the photon is energetic and collinear with the muon. Therefore, we reject electrons if there is a reconstructed muon within $\Delta R$=0.1. This rejection could be accomplished more elegantly by ensuring that the electron track and the

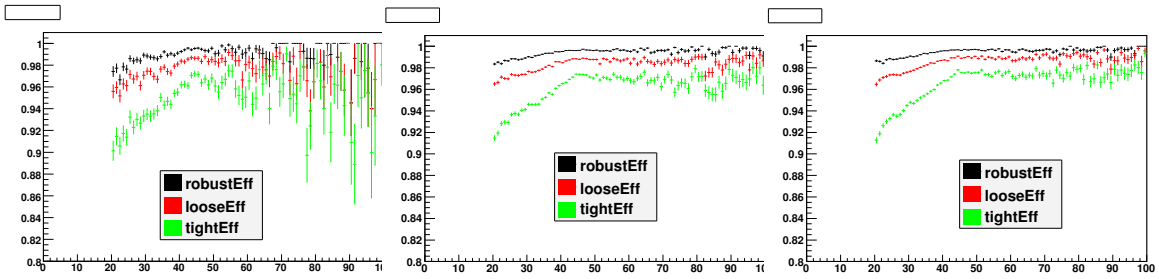



Figure 4: Efficiency for the robust, loose, and tight selections as a function of $P_T$ for electrons in $Z \rightarrow ee$. Left plot: in the startup scenario; middle plot: 10 pb$^{-1}$ calibrations; Right plot: 100 pb$^{-1}$ calibrations. See text for the precise definition of efficiency.

muon track are different, *i.e.*, that they do not share too many hits. However, the rough $\Delta R$ requirement is also very effective, and no efficiency is lost.

Photon conversions are an important source of fake electrons in QCD events. We have found that about 35% of isolated high $P_T$ electrons in QCD events, passing the tight selection, are from photon conversions[12]. We have also found that a moderate impact parameter cut, *e.g.*, $|d_0| < 250\ \mu m$, rejects a significant fractions of the conversion background. We are now developing an algorithm to reject conversions by finding the partner track[13]. Since this algorithm is still under development, we do not apply it to our analysis. Based on our study, we had previously proposed to include a $|d_0| < 250\ \mu m$ requirement to our electron selection[14]. However, in the first phase of data taking the impact parameter resolution is not very good, see Figure 5. For the purpose of this note, we place an impact parameter requirement $|d_0|$ <400 $\mu m$, which should already be safe with 10 pb$^{-1}$ calibrations.

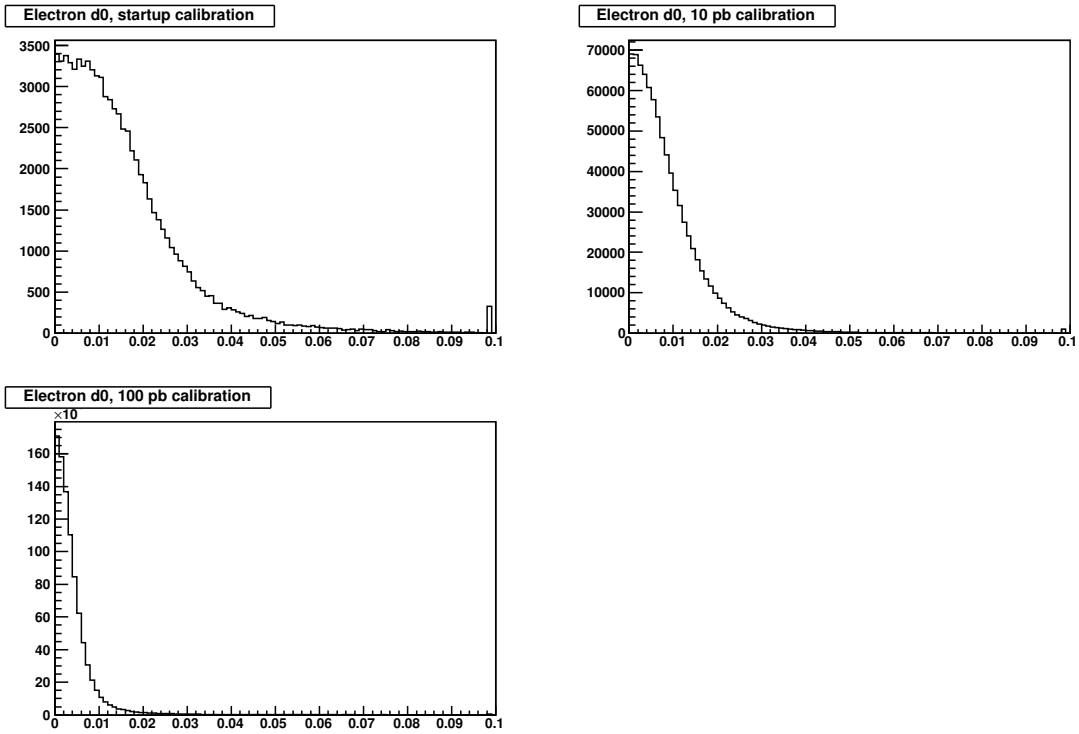



Figure 5: Impact parameter (cm) for electrons passing the tight selection and truth matched to $Z \rightarrow ee$ decays. Top left: startup calibrations; top right: 10 pb$^{-1}$ calibrations; bottom left: 100 pb$^{-1}$ calibrations.

To summarize, our base electron selection is

- Tight from the egamma POG.
- Reject electrons within $\Delta R$=0.1 of any muon candidate.

- $|d_0| < 400\ \mu m$

Just as for muons, isolation is an important handle against fake electrons in QCD events. We will discuss isolation in Section 6.4.

## 6.3 Missing Transverse Energy

In the absence of $b$-tagging, Missing Transverse Energy ($\not\!\!E_T$) is the main handle against the Drell-Yan backgrounds. In our base event selection, we do not correct the $\not\!\!E_T$ for the jet energy scale (JES). This is mostly because the correction does not help in rejecting the Drell-Yan background, see Figure 6.

We note that the available JES-based correction to the $\not\!\!E_T$ is inconsistent in its treatment of the out-of-cone correction[15]. Basically the issue is that the energy that ends up out of the jet cone, *e.g.*, because of the **B**-field bend, is double counted in the $\not\!\!E_T$ calculation. This energy is counted as unclustered energy, and also *put back* into the jet cone on a statistical basis by the JES correction. It is not clear how big of an effect this is, but it could potentially be very important for the low $E_T$ jets that we are dealing with.

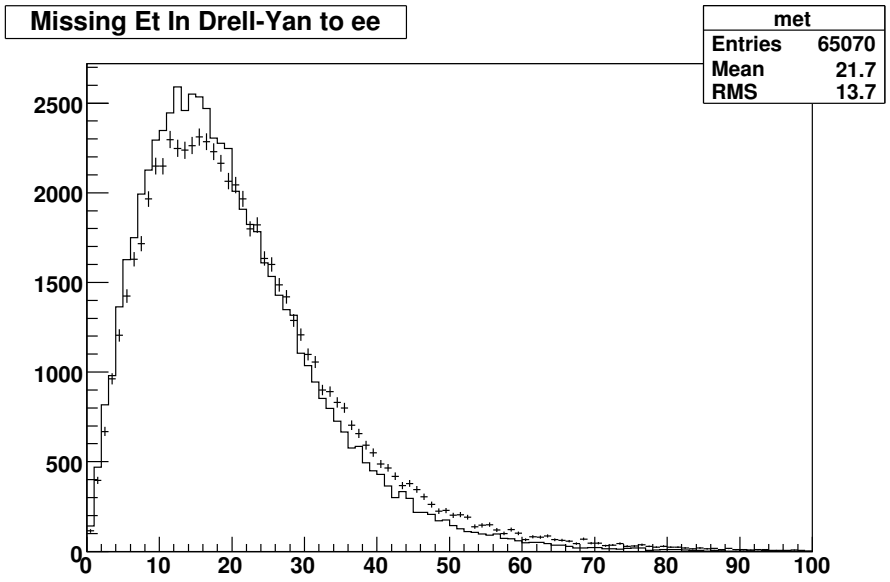



Figure 6: Missing Transverse Energy in MC Drell Yan $\rightarrow ee$ events with $\geq 1$ additional jets of uncorrected $P_T > 15$ GeV, with (points) and without (histogram) the jet energy scale correction. This plot was made on a 1_5 Pythia Monte Carlo sample.

Unlike the default TQAF, we do not correct the $\not\!\!E_T$ for all muons in the event. Instead, we only correct the $\not\!\!E_T$ for the isolated muons that make up our dilepton event solution. The reason for this is that the standard muon $\not\!\!E_T$ corrections are not intended to be used for muons in jets[16]. However, we have verified that these two alternative definitions of muon-corrected $\not\!\!E_T$ make a negligible difference to our analysis. In the future these effects may become more important since in more recent versions of CMSSW the defult muon collection will also include TrackerMuons, which by design are less pure when used out of the box.

The $\not\!\!E_T$ requirement is very important to reject Drell-Yan. As discussed in Section 2, the Drell-Yan background is most important in the same-flavor channels. Thus, our base selection has $\not\!\!E_T$ selection requirements for $ee$ and $\mu\mu$ that are different than those for $e\mu$.

### 6.3.1 Missing Transverse Energy for $ee$ and $\mu\mu$

For the $ee$ and $\mu\mu$ channels, we picked $\not\!\!E_T$ requirements that are not too stringent and that are targeted against the Drell-Yan background. In Drell-Yan events the $\not\!\!E_T$ is dominantly from misemasurements of the hadronic part of the event. Because of momentum conservation, the net hadronic recoil is back-to-back with the dilepton $P_T$ in the transverse plane. Without JES corrections, the hadronic recoil tends to be undermeasured, thus the $\not\!\!E_T$ tends to


Figure 7: The angle $\alpha$ defined in the text.

be back-to-back with the dilepton $P_T$. We define an angle $\alpha$ between the missing $E_T$ vector and the transverse component of the dilepton momentum vector ($P_T^{\ell\ell}$) according to Figure 7. Our $\not\!\!E_T$requirements are then

- $\not\!\!E_T >$ 30 GeV
- $\not\!\!E_T > 0.6 \cdot P_T^{\ell\ell}$ OR $\alpha > 0.25$

The 2nd requirement can be visualized in Figure 8. Note that this requirement is quite efficient not just for $t\bar{t}$, but also for other interesting dilepton physics processes such as $WW$.

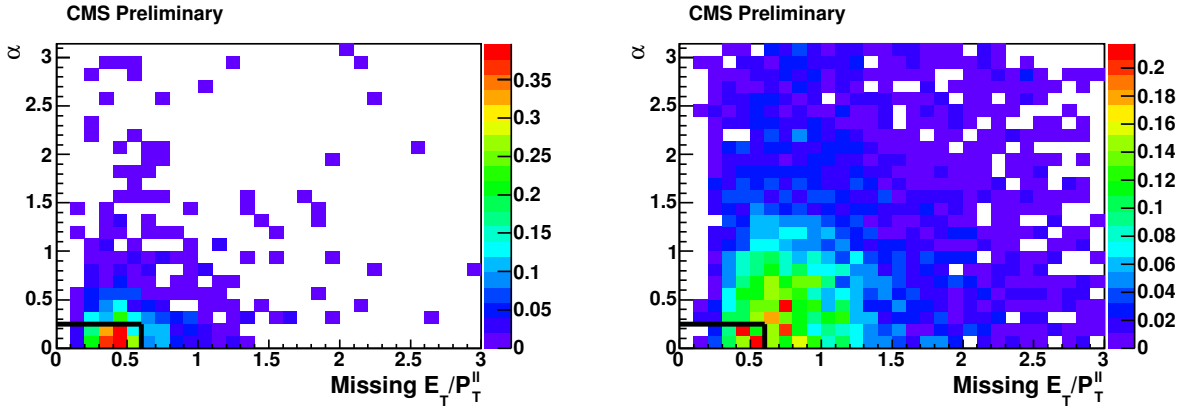



Figure 8: Scatter plot of MET/$P_T^{\ell\ell}$ vs. the angle $\alpha$ defined in Figure 7 for Drell-Yan $\rightarrow ee$ (left) and $t\bar{t} \rightarrow ee$ (right) normalized to 10 pb$^{-1}$. This is for events with MET $>$ 30 GeV and $\geq$ 2 jets. Events in the box at the botton-left-hand corner of the scatter plot are rejected by our requirement.

For events with at least two jets ($P_T > 30$ corrected and $|\eta| < 2.4$), these requirements reject 81% of the DY$\rightarrow ee$ background and are 86% efficient for $t\bar{t} \rightarrow ee$. Alternatively, one could simply cut tighter on $\not\!\!E_T$. For example, in Reference [1] the proposed $\not\!\!E_T$requirement is simply $\not\!\!E_T >$ 50 GeV, where the requirement is on the JES- and muon-corrected $\not\!\!E_T$.

### 6.3.2 Missing Transverse Energy for $e\mu$

In the case of $e\mu$, one could in principle get away without a $\not\!\!E_T$cut, since in this channel the most important backgrounds ($WW$ and $W$+jets) have $\not\!\!E_T$similar to the $t\bar{t}$ signal. We do apply a $\not\!\!E_T >$ 20 GeV requirement to reduce a bit the Drell-Yan $\rightarrow \tau\tau$ background, and also since such a requirement helps somewhat in reducing purely QCD backgrounds, see Section 8.

## 6.4 Isolation

We require both leptons to be isolated. For muons we define a variable

$S = \sum \text{track } P_T + \sum \text{HCAL } E_T + \sum \text{ECAL } E_T$

where the sums are in a cone of $\Delta R = 0.3$ taking care of not counting the muon track $P_T$ and the muon energy deposition. These sums are taken straight out of the muon object.

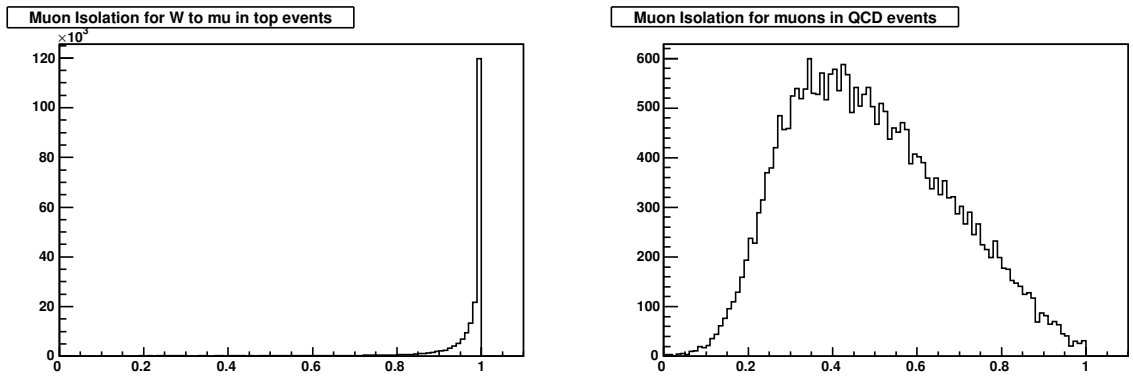



Figure 9: The isolation variable defined in the text for muons from $W$ decays in top events and for muons of in the ppMuPt20-15 QCD sample described in Section 8. These distributions were obtained after the skim selection, for muons that passed the selection criteria of Section 6.1.

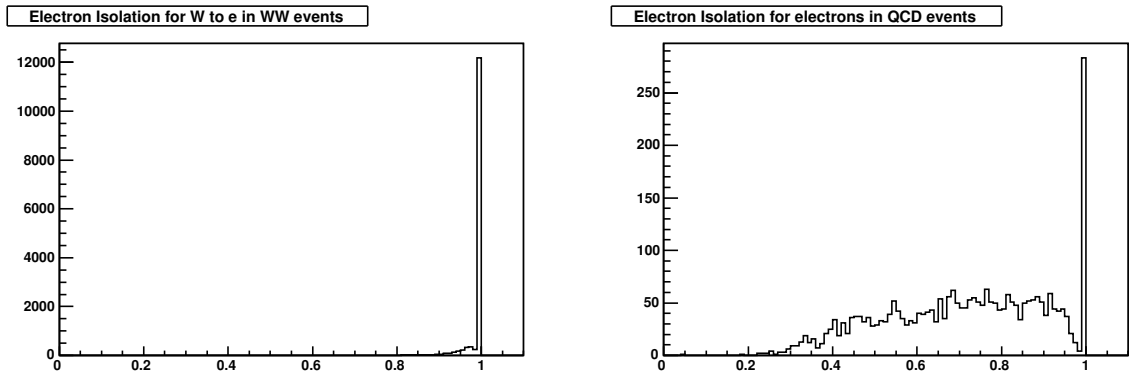



Figure 10: The isolation variable defined in the text for electrons from $W$ decays in $WW$ events and for electrons of in the ppMuPt20-15 QCD sample described in Section 8. These distributions were obtained after the skim selection, for electrons that passed the tight selection criteria of Section 6.2.

For electrons we define (for now) a simpler variable

$S = \sum \text{track } P_T$

where once again the sum is in a cone of $\Delta R = 0.3$, and the electron track is not counted. We calculate this sum ourselves, rather than using the one provided by the TQAF. This is because the TQAF sum does not apply any track quality cuts on the tracks in the sum. Our sum is constructed from tracks in the cone with $P_T > 1$ GeV, excluding the electron track. Our track quality requirements are $|d_0| < 1$ mm, at least 7 hits on the track, and $\Delta z$ between the electron track and the track in question $< 5$ mm.

We do not use the TQAF calorimeter sum from the TopElectron object. This quantity is at the moment calculated in an inconsistent way and can be negative. We will switch to the variables provided by the Egamma POG when they become available.

We then require, for both electrons and muons:

$$Iso \equiv \frac{P_T^{\text{lepton}}}{P_T^{\text{lepton}} + S} > 0.92$$

The $Iso$ variable defined above has been used in CDF analyses; it spans the interval 0 to 1. Note that a cut on $Iso$ is equivalent to a cut on relative isolation, $S/P_T^{\text{lepton}}$. We choose to cut on relative isolation instead of absolute isolation because a cut on relative isolation is by construction tightest where most of the QCD backgrounds is, *i.e.*, near the $P_T^{\text{lepton}}$ threshold.

Distributions of $Iso$ for muons in top events and in the ppMuPt20-15 QCD sample described in Section 8 are

shown In Figure 9. On top events, the efficiency of the requirement is 87%. Muons in QCD events are rejected by a facor of about 85.

Because our top sample is filtered through the HLT, which applies its own isolation requirement on electrons, we show the unbiased isolation efficiency distribution for electrons in $WW$ events in Figure 10. On $WW$ events, the efficiency is 96%. Electrons in QCD events are rejected by a factor of about 10.

Clearly the $Iso$ requirement for electrons could be made tighter by adding information on calorimeter energy. This will be an important future improvement to our base selection.

### 6.5 $Z$ veto

We reject all events where there are two leptons of the same flavor passing our base selection with invariant mass between 76 and 106 GeV. We allow one of the leptons used to reconstruct the $Z$ to be non-isolated.

The veto introduces an inefficiency of 25% on the $t\bar{t} \to ee$ or $\mu\mu$ signal. This veto is necessary because of the high cross-section for $Z$ production and the imperfect $\not\!\!E_T$resolution.

### 6.6 Jet Selection

We need to select jets for jet counting. We use caloJets reconstructed with the $\Delta R = 0.5$ iterativeCone algorithm.

As mentioned in Section 5, we do not use the TQAF facility to avoid the double counting of jets and electrons. The reason for this is that we want to make sure that our jet and electron selections are totally synchronized. We prefer to remove jets that are matched to electrons after the electrons have been fully identified and not before. Just as in the TQAF, the removal is done by matching jets and electrons in $\Delta R$.

Our base selection is to count jets with corrected $P_T > 30$ GeV and $|\eta| < 2.4$. Note that our base selection used to have uncorrected $P_T > 15$ GeV and $|\eta| < 3$; we recently switched to the more standard top group selection [1] described here. Some of the studies described in this note were done with the earlier jet selection. In many cases the details of the event selection do not matter, and therefore in some cases the studies were not repeated with the newer selection.

### 6.7 Summary of Event Selections

We summarize in Table 4 the changes in the number of expected events in the signal and control (events with 0 and 1 jets) samples as the cuts are applied to events in all the dilepton channels combined. Up to this point the events in separate dilepton channels have been treated independently. We note that there are cases in which several dilepton candidates pass the selections. For the final event selection in which the events in all channels are combined we apply a disambiguation requirement. As a result only one candidate per event is accepted: we count only the $e\mu$ candidates in events with $ee$ or $\mu\mu$ candidates passing the event selections; we accept only the first candidate if multiple candidates pass selections in the same dilepton channel. The effect of this requirement is negligible for our analysis.

To give a further insight on the selection flow in $ee$, $\mu\mu$, and $e\mu$ we show the corresponding number of expected events in Tables 5, 6, and 7. The number of events resulting from $t\bar{t}$ where each lepton is found to originate from a $W$ decay based on the MC information are shown separately from all other cases of $t\bar{t}$ decays. This separation is helpful to see that the applied requirements are efficient on events with the genuine expected signal.

## 7 Results

### 7.1 Event counts with the base selection

In Figures 11 we show the expected event count as a function of $N_{jets}$ for the base selection. The number of expected events with 0, 1, and $\geq 2$ jets corresponding to the control and signal bins are summarized in Table 7.1. These distributions have been obtained using the Chowder $t\bar{t}$, $W+$ jets and Drell Yan data samples with 10 pb$^{-1}$ calibrations. The $WW$, $WZ$, and $ZZ$ data samples have 100 pb$^{-1}$ calibrations. The purely QCD backgrounds will be addressed in Section 8.

Some interesting points to note:

| | 0, 1 jets | | | | $\geq 2$ jets | | | |
|---|---|---|---|---|---|---|---|---|
| Source | $t\bar{t}$ | $VV$ | $W$+jets | DY+jets | $t\bar{t}$ | $VV$ | $W$+jets | DY+jets |
| After skim | 103.27 | 42.62 | 637.11 | 12234.00 | 520.54 | 16.80 | 141.58 | 765.32 |
| AND lepton ID | 59.62 | 34.12 | 55.70 | 11441.89 | 261.29 | 12.27 | 20.13 | 677.37 |
| AND isolation | 44.42 | 30.18 | 8.01 | 10173.29 | 133.60 | 9.81 | 1.04 | 537.26 |
| AND $Z$-veto | 39.22 | 15.76 | 7.05 | 884.50 | 118.49 | 2.46 | 0.96 | 58.67 |
| AND $\not\!\!E_T$ | 33.62 | 11.56 | 5.51 | 26.86 | 101.63 | 1.70 | 0.67 | 11.42 |
| No lepton ID | 38.09 | 13.04 | 50.58 | 51.47 | 119.96 | 2.11 | 7.19 | 15.64 |
| No isolation | 45.00 | 13.21 | 36.96 | 33.61 | 195.50 | 2.48 | 13.70 | 17.19 |
| No $Z$-veto | 37.70 | 17.80 | 5.97 | 242.41 | 113.47 | 3.98 | 0.74 | 97.45 |
| All (no duplicates) | 33.59 | 11.54 | 5.51 | 26.86 | 101.50 | 1.68 | 0.67 | 11.41 |

Table 4: The expected numbers of events with opposite sign lepton pairs in the signal (events with at least two jets) and control samples (events with 0 and 1 jets) at different stages of event selection. The numbers shown are for all the dilepton channels combined. Here $VV$ corresponds to the combination of $WW$, $WZ$, and $ZZ$ production processes. The first four rows correspond to the number of events expected to pass the selection in the order from the loosest (all events collected in the sample) to the tightest (passing all requirements). The following three rows correspond to the number of events expected to pass all except the mentioned requirement. The last row corresponds to events with multiple dilepton candidates removed following the prodecure detailed in text.

| Source | $t\bar{t} \to ee$ | other $t\bar{t}$ | $VV$ | $W$+jets | DY+jets |
|---|---|---|---|---|---|
| After skim | 26.78 | 111.24 | 6.19 | 52.18 | 289.43 |
| AND lepton ID | 24.13 | 15.40 | 4.15 | 2.95 | 235.83 |
| AND isolation | 21.71 | 6.88 | 3.78 | 0.51 | 210.36 |
| AND $Z$-veto | 17.16 | 5.29 | 0.58 | 0.43 | 19.84 |
| AND $\not\!\!E_T$ | 13.42 | 4.18 | 0.32 | 0.20 | 3.81 |
| No lepton ID | 14.76 | 8.76 | 0.45 | 2.39 | 4.86 |
| No isolation | 15.00 | 8.88 | 0.38 | 1.39 | 4.53 |
| No $Z$-veto | 17.04 | 5.42 | 1.26 | 0.27 | 38.61 |
| All (no duplicates) | 13.42 | 4.15 | 0.31 | 0.20 | 3.81 |

Table 5: The expected numbers of events with two opposite sign electrons and two or more jets at different stages of event selection. The first column with numbers ($t\bar{t} \to ee$) corresponds to the simulated events in which both $W$ bosons are required to decay to electrons. The second column (other $t\bar{t}$) corresponds to the remaining cases of top quark decays. The remaining notations in the table are the same as in Table 4.

- As anticipated, the $ee$ and $\mu\mu$ channels suffer from large Drell Yan backgrounds.
- In all channels, but particularly in the $e\mu$ channel, events with $N_{jets} \geq 2$ are dominantly $t\bar{t}$.
- The expected event yields are such that the statistical uncertainty on a cross-section measurement in 10 pb$^{-1}$ will be of order 10%.
- The $W+$ jets background is small. It consists mostly of events with a fake electron. The anticipated addition of electron calorimeter isolation and conversion rejection will reduce this background further.
- The diboson backgrounds are small.
- There is a small background in the $e\mu$ channel from Drell Yan $\to \mu\mu$. These are events where the muon comes from the $\gamma^*/Z^{(*)}$, and the electron is fake.
- For the $t\bar{t}$ sample, 87% of the events with $\geq 2$ jets are from the dilepton decay modes ($t\bar{t} \to ee, \mu\mu, e\mu$). The remainder are almost entirely lepton + $\tau$ (12%).

Note that effect of multiple interactions in the same beam-beam collision (pile-up) have not been simulated or taken into account. It is forseeable that this effect will be negligible in the first $10$ pb$^{-1}$ of data considered for this analysis.

| Source | $t\bar{t} \to \mu\mu$ | other $t\bar{t}$ | $VV$ | $W$+jets | DY+jets |
|---|---|---|---|---|---|
| After skim | 44.87 | 57.36 | 6.51 | 11.99 | 447.15 |
| AND lepton ID | 43.99 | 52.45 | 6.28 | 8.23 | 437.06 |
| AND isolation | 33.59 | 4.52 | 4.71 | 0.00 | 324.99 |
| AND $Z$-veto | 26.23 | 3.48 | 0.83 | 0.00 | 37.02 |
| AND $\not\!\!E_T$ | 20.54 | 2.70 | 0.48 | 0.00 | 6.62 |
| No lepton ID | 20.99 | 2.87 | 0.50 | 0.12 | 7.13 |
| No isolation | 27.27 | 31.49 | 0.78 | 4.34 | 10.71 |
| No $Z$-veto | 26.19 | 3.50 | 1.58 | 0.00 | 57.76 |
| All (no duplicates) | 20.54 | 2.64 | 0.46 | 0.0 | 6.61 |

Table 6: The expected numbers of events with two opposite sign muons and two or more jets at different stages of event selection. The first column with numbers ($t\bar{t} \to \mu\mu$) corresponds to the simulated events in which both $W$ bosons are required to decay to muons. The second column (other $t\bar{t}$) corresponds to the remaining cases of top quark decays. After final selections these events are dominated by cases where one $W$ boson decayed to a muon and the other one decayed to a $\tau$ which subsequently decayed to a muon. The notations are the same as in Table 5.

| Source | $t\bar{t} \to e\mu$ | other $t\bar{t}$ | $VV$ | $W$+jets | DY+jets |
|---|---|---|---|---|---|
| After skim | 70.93 | 209.35 | 4.10 | 77.41 | 28.72 |
| AND lepton ID | 66.62 | 58.68 | 1.84 | 8.95 | 4.48 |
| AND isolation | 54.66 | 12.25 | 1.32 | 0.53 | 1.92 |
| AND $Z$-veto | 54.19 | 12.14 | 1.06 | 0.53 | 1.81 |
| AND $\not\!\!E_T$ | 49.68 | 11.11 | 0.91 | 0.47 | 0.99 |
| No lepton ID | 52.68 | 19.90 | 1.16 | 4.68 | 3.65 |
| No isolation | 60.75 | 52.10 | 1.31 | 7.97 | 1.95 |
| No $Z$-veto | 50.11 | 11.21 | 1.15 | 0.47 | 1.09 |
| All (no duplicates) | 49.68 | 11.07 | 0.91 | 0.47 | 0.99 |

Table 7: The expected numbers of events with an opposite sign electron-muon pair and two or more jets at different stages of event selection. The first column with numbers ($t\bar{t} \to e\mu$) corresponds to the simulated events in which one $W$ boson is required to decay to a muon and the other one is required to decay to an electron. The second column (other $t\bar{t}$) corresponds to the remaining cases of top quark decays. The notations are the same as in Table 5.

## 7.2 Study of impact electron ID

To test the impact of electron ID, we focus on $t\bar{t}$ and $W+$ jets events. The latter are dominated by $W \to \mu\nu$ with a fake electron. In Figure 12 we show the $N_{jets}$ count in the $e\mu$ channel for the base selection (tight electron ID), and for less stringent electron selections (loose and robust electron ID). These distributions were obtained with the same calibrations as those in Figure 11.

In Figure 12 we see that the $W \to \mu\nu +$ fale electron background (in grey) is considerably reduced as the electron identification requirements are tightened, while the signal efficiency does not change very much. This background is quite small in the $\geq 2$ jets bins, but is one of the dominant contributions in the 0 and 1 jet bins. As discussed in Section 2, we want to use the 0 and 1 jet bins as control samples for the background predictions. We anticipate that accurately predicting the fake electron background will be challenging, therefore we prefer to keep it as small as possible everywhere. Furthermore, we are also interested in pursuing $WW$ physics in the 0 jet bin from the same platform. Thus, we choose the tight selection from the egamma POG for this analysis, since the efficiency is still good and the background is minimized.

## 7.3 Kinematical Distributions

In this Section we show some of the most interesting kinematical distribuitions.

The $\not\!\!E_T$ distribution is shown in Figure 13 for events with $\geq 2$ jets and a cut on $\not\!\!E_T$ value relaxed (the topological cut $\not\!\!E_T > 0.6 \cdot P_T^{\ell\ell}$ descibed in Section 6.3 is applied). It can be seen that a substantial fraction of DY+jets background is removed by the additional $\not\!\!E_T > 30$ GeV cut in the $ee$ and $\mu\mu$ channels.

**Need to stick some of the standard plots tin this section. No big deal.**

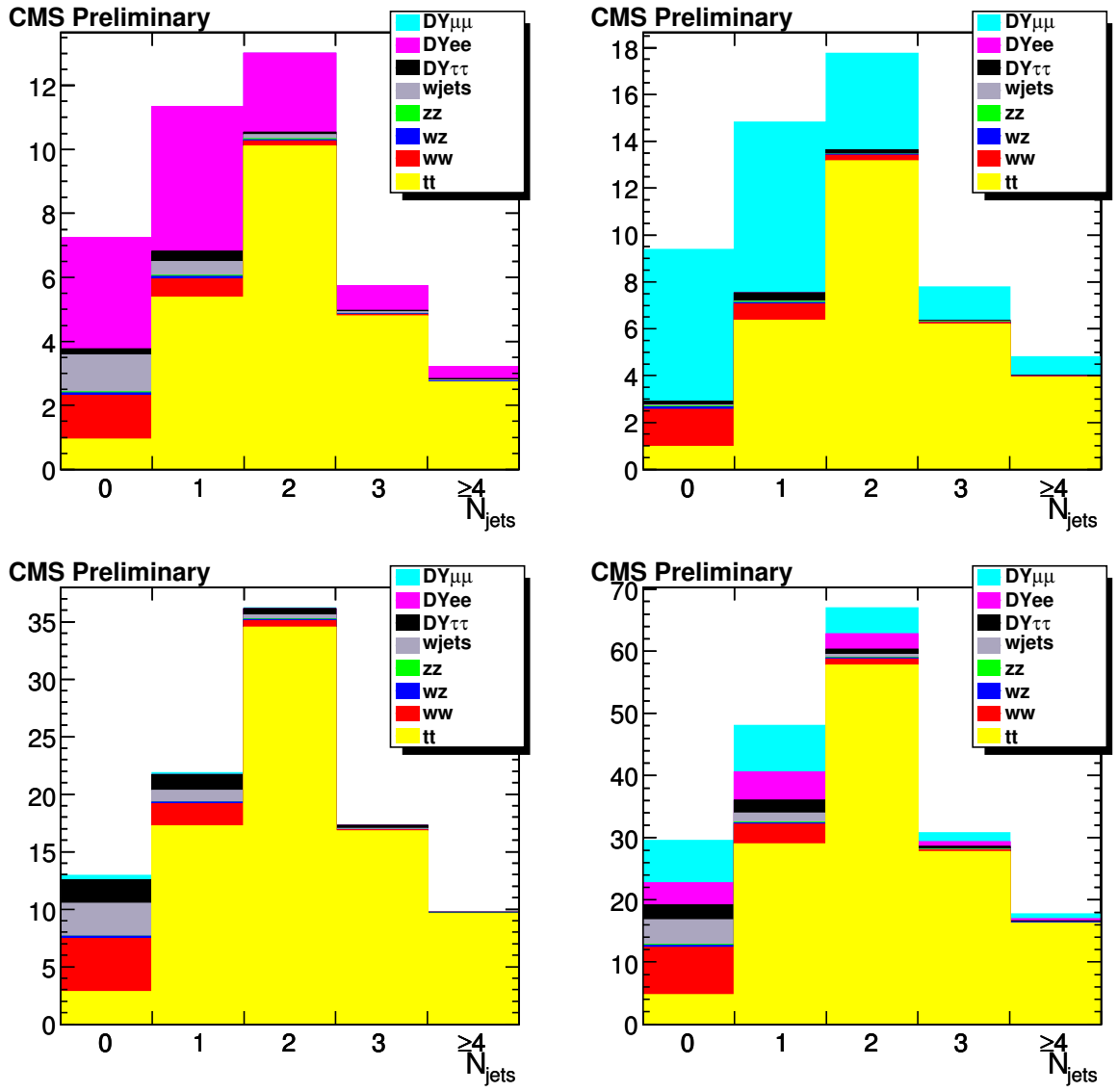



Figure 11: The expected number of dilepton events in 10 pb$^{-1}$ as a function of jet multiplicity from various sources, with the selection of Section 6. Top left: $ee$; top right: $\mu\mu$; bottom left: $e\mu$; bottom right: all combined. The color coding is the following: yellow=$t\bar{t}$, red = $WW$, dark blue = $WZ$, green = $ZZ$, grey = $W$+jets, black = DY$\rightarrow \tau\tau$, magenta = DY$\rightarrow ee$, cyan = DY$\rightarrow \mu\mu$.

# 8 QCD Backgrounds

By QCD backgrounds we denote those backgrounds where neither of the two leptons come from weak boson decays. We call these leptons fake, even if some of them are actually real leptons from decays of heavy quarks.

In Section 12 we will investigate methods to measure these backgrounds, as well as the $W$+jets background, in a data-driven way. In this Section, on the other hand, we describe the use of Monte Carlo to estimate the expected QCD background in our data. This is important because we want to make sure that our carefully crafted event selection is not swamped by QCD backgrounds.

The available QCD Monte Carlo samples have limited statistics. They are also sensitive to significant theoretical uncertainties on $b\bar{b}$ production, and rely on the simulation for the modeling of instrumental fakes. Thus they can only provide a ballpark estimate of the real QCD backgrounds that we will face in CMS.

Here we will try as much as possible to use specialized Monte Carlo samples to directly estimate the QCD backgrounds. Others [1] have done similar studies on generic QCD samples using a factorization approach.

| Source | $ee$ | | $\mu\mu$ | | $e\mu$ | | all | |
|---|---|---|---|---|---|---|---|---|
| | $0,1$ jets | $\geq 2$ jets | $0,1$ jets | $\geq 2$ jets | $0,1$ jets | $\geq 2$ jets | $0,1$ jets | $\geq 2$ jets |
| $WW$ | 1.9 | 0.2 | 2.2 | 0.3 | 6.6 | 0.8 | 10.7 | 1.3 |
| $WZ$ | 0.2 | 0.1 | 0.2 | 0.1 | 0.3 | 0.1 | 0.7 | 0.3 |
| $ZZ$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.03 | 0.02 | 0.2 | 0.1 |
| $W$+jets | 1.6 | 0.2 | 0.1 | 0.0 | 3.8 | 0.5 | 5.5 | 0.7 |
| DY$\rightarrow \tau\tau$ | 0.5 | 0.2 | 0.5 | 0.2 | 3.4 | 0.9 | 4.4 | 1.3 |
| DY$\rightarrow ee$ | 8.0 | 3.6 | 0.0 | 0.0 | 0.0 | 0.0 | 8.0 | 3.6 |
| DY$\rightarrow \mu\mu$ | 0.0 | 0.0 | 13.8 | 6.4 | 0.6 | 0.1 | 14.4 | 6.5 |
| Total non $t\bar{t}$ | 12.3 | 4.3 | 16.9 | 7.1 | 14.8 | 2.4 | 43.9 | 13.8 |
| $t\bar{t}$ | 6.3 | 17.6 | 7.3 | 23.2 | 20.0 | 60.7 | 33.6 | 101.5 |
| Total | 18.6 | 21.9 | 24.2 | 30.3 | 34.8 | 63.2 | 77.6 | 115.4 |

Table 8: Expected number of events after all selections are applied

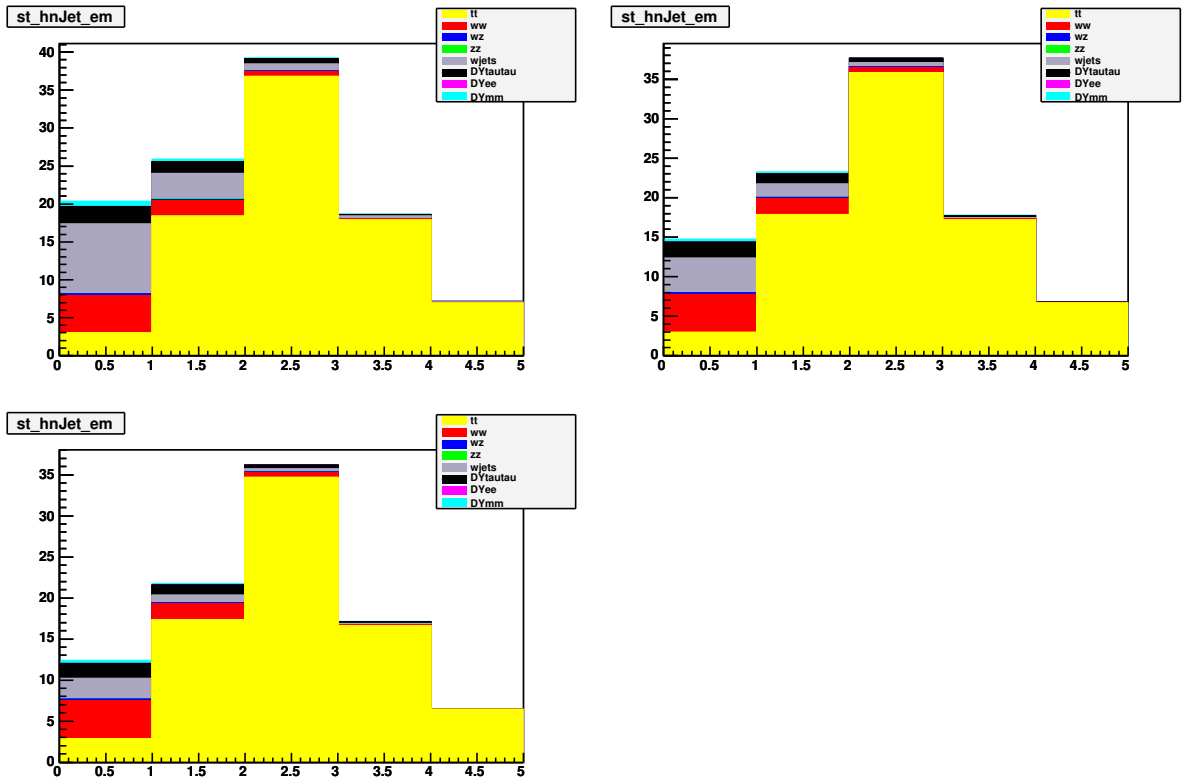



Figure 12: The expected number of $e\mu$ events in 10 pb$^{-1}$ as a function of jet multiplicity. Top left: base selection but with tight electron ID; Top right: base selection but with loose electron ID; Bottom left: base selection (tight electron ID). The color scheme is the same as in Figure 11.

## 8.1 QCD background in $e\mu$ and $\mu\mu$

As mentioned in Section 6.1, after the selection requirements the rate of muons in QCD Monte Carlo events is dominated by $b \rightarrow \mu$ and $c \rightarrow \mu$. Thus we use the ppMuPt20-15 sample [18] to estimate the $\mu\mu$ and $e\mu$ backgrounds. This is a Pythia QCD sample with $\hat{P_T} > 20$ GeV filtered at the generator level for the existence of a muon of $P_T > 15$ GeV.

We run the standard selection on this sample. In the $\mu\mu$ channel we find no candidates passing the base selection, regardless of jet multiplicity. In the $e\mu$ channel there is ony one candidate ($e\mu + 1$ jet) that survived. Since the integrated luminosity of this sample is 8.7 pb$^{-1}$, we can conclude that in 10 pb$^{-1}$ Pythia predicts a small QCD background in the $\mu\mu$ and $e\mu$ channels. Note that, as mentioned in Section 6.4, there is room for improvement in electron isolation, thus the QCD background in $e\mu$ can be reduced further.

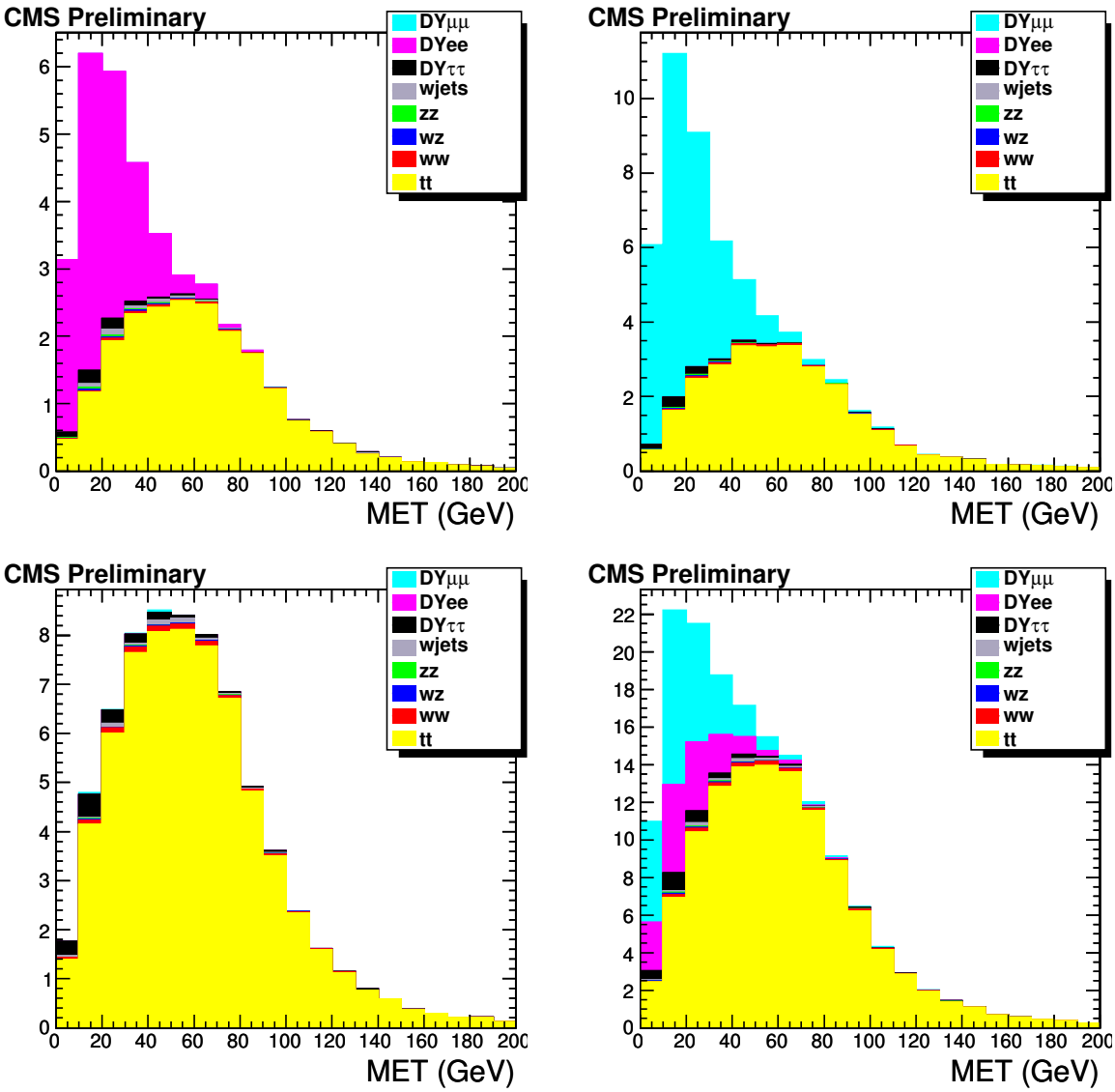



Figure 13: Distribution of $\not\!\!E_T$ in $ee$ (top left), $\mu\mu$ (top right), $e\mu$ (bottom left), and all channels combined (bottom right) events. The events are passing all requirements except for $\not\!\!E_T > 30(20)$ GeV in $ee$, $\mu\mu$ ($e\mu$) cases.

In Section 6.3 we stated that a moderate $\not\!\!E_T > 20$ GeV requirement can be used to reduce the QCD background in the $e\mu$ channel. If we remove the $\not\!\!E_T$ requirement on the $e\mu$ channel we find that the number of events surviving the base selection in the ppMuPt20-15 sample doubles, *i.e.*, it goes from one to two. To get better statistics, we show in Figure 14 the $\not\!\!E_T$ distribution of $e\mu$ events from the ppMuPt20-15 sample with isolation removed from the selection. A $\not\!\!E_T > 20$ GeV requirement would reject 46% of these events. This requirement has an efficiency of 95% on $t\bar{t} \to e\mu$ events.

If the Pythia prediction is correct, the $\not\!\!E_T$ requirement is not really needed in the $e\mu$ sample. For the moment we keep this requirement with the understanding that the situation will have to be re-examined with real data.

## 8.2 QCD background in $ee$

Next, we address the QCD background in $ee$. We used the EM QCD Pythia enriched samples listed in Table 8.2 with their nominal integrated luminosities

These samples were filtered at the generator level to simulate only events with jets rich in EM content, and thus likely to contain fake electrons. This is a very difficult thing to do, and studies [19] have shown that the filtering captures only about one half of fake electrons in QCD events. Since we are interested in $ee$ background events, the

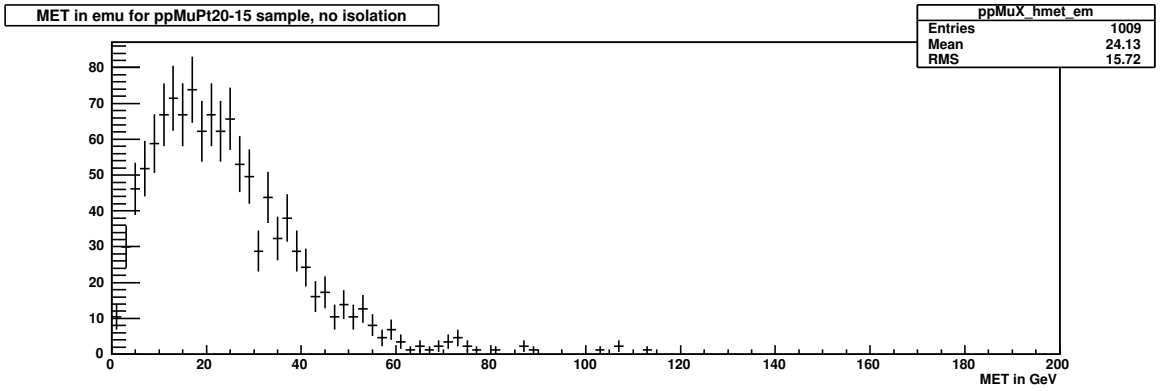



Figure 14: Missing Transverse Energy in MC ppMuPt20-15 $e\mu$ events passing all base selection requirements, except MET and isolation. This distribution is integrated over all $N_{jets}$ bins.

| | |
|---|---|
| QCDJetsEnriched_Pt_25_50/CMSSW_1_6_7-CSA07-1192837439/RECO | 0.23 pb$^{-1}$ |
| QCDJetsEnriched_Pt_50_170/CMSSW_1_6_7-CSA07-1193937628/RECO | 0.53 pb$^{-1}$ |
| QCDJetsEnriched_Pt_170_up/CMSSW_1_6_7-CSA07-1192835861/RECO | 23.9 pb$^{-1}$ |

Table 9: The EM enriched QCD samples, and their nominal integrated luminosities. The first one is generated wit $25 < \hat{P}_T < 50$ GeV, the second one with $50 < \hat{P}_T < 170$ GeV, and the third with $\hat{P}_T > 170$ GeV.

filter has two chances to "trigger"; thus we rescale the luminosities in Table 8.2 by $1 - 0.5^2 = 0.75$.

Running our analysis on the three data sets of Table 8.2, we find no $ee$ candidates passing our selection, regardless of jet multiplicity. Fot the high statistics $\hat{P}_T > 170$ GeV data set, this translates into a 90% CL limit of $< 1.3$ events in 10 pb$^{-1}$. The limits for the two lower $\hat{P}_T$ sets are not useful because of the low statistics.

Next, we repeat the analysis on the two low $\hat{P}_T$ data sets with the electron isolation and trigger requirements removed. From the discussion of Section 6.4, our current electron isolation requirements reduce QCD backgrounds by about a factor of 10 per electron. We assume that we can factorize the efficiencies of electron isolation and all other selection requirements. Then, the effective integrated luminosities of these samples, when analyzed with no isolation and trigger[3)] requirements, increase by $\approx 10 \times 10 = 100$.

Even without the isolation/trigger requirements, no events from the two low $\hat{P}_T$ data sets pass the selection. Assuming the factor of 100 from the factorization assumption, this results in 90% CL limits of $< 1.3$ and $< 0.6$ events in 10 pb$^{-1}$ for the $25 < \hat{P}_T < 50$ GeV sample and the $50 < \hat{P}_T < 170$ GeV sample, respectively.

In conclusion, the EM enriched sample indicate that the $ee$ QCD background is small: less than a couple of events integrated over $N_{jets}$. This small background should get even smaller when adding calorimeter isolation and conversion removal to the base selection.

# 9 Calibration Issues

## 9.1 Calibration/Alignment Scenarios

To test our sensitivity to calibration issues, we repeat our analysis using the 100 pb$^{-1}$ calibration constants for the soup datasets. The changes in the number of accepted $t\bar{t}$ events with $N_{jets} \geq 2$ is summarized in Table 9.1.

| $ee$ mode | $e\mu$ mode | $\mu\mu$ mode | all combined |
|---|---|---|---|
| $+10.4\%$ | $+3.5\%$ | $-0.5\%$ | $+3.8\%$ |

Table 10: Changes in $t\bar{t}$ acceptance in going from 10 pb$^{-1}$ to 100 pb$^{-1}$ calibrations. There is a systematic uncertainties at the percent level due to book-keeping complications.

From Table 9.1 we conclude that with 10 pb$^{-1}$ calibrations there is an efficiency loss of order 5% per electron

[3)] We removed the electron trigger requirement since isolation is required in the electron trigger at the HLT.

in the final state with respect to the 100 pb$^{-1}$ calibrations. Note that the efficiency of the tight egamma electron selection with respect to the offline reconstruction cannot explain this difference, see Figure 4. We suspect that this inefficiency is in the offline reconstruction. This would not be totally surprising, since one of the most delicate parts of the electron reconstuction is the pixel seeding, which is quite sensitive to alignment issues. (Note tha miscalibrations were not applied at the HLT.).

### 9.2 Jet Energy Scale Uncertainty

One of the major systematics on the acceptance is the uncertainty on the jet energy scale. We repeat the analysis by shifting the jet thresholds for jet counting by $\pm 10\%$. We find a change in $t\bar{t}$ events with $N_{jets} \geq 2$ of $\pm 5\%$. Note that this does not include the uncertainty due to the correlated miscalibration in $\not\!\!E_T$.

## 10 Lepton Efficiency Determination

We will measure the lepton efficiency from data using the standard tag-and-probe method [17]. With a data sample of 10 pb$^{-1}$ we should have enough $Z$s in our sample to be able to estimate the lepton efficiencies to the percent level.

The isolation efficiency for lepton in top events is different than that for leptons in $Z$ events, since $t\bar{t}$ events tend to have more hadronic activity. To estimate this efficiency and its related uncertainty, we will first compare the efficiency in Monte Carlo $Z$ events with data $Z$ events. This will give us a baseline understanding of the ability of the Monte Carlo to reproduce the isolation efficiency in relatively quiet events. Next, we will use the Monte Carlo to estimate the difference in lepton isolation efficiency between $t\bar{t}$ and $Z$ events. We will use different $t\bar{t}$ Monte Carlo models to assign a systematic uncertainty. We guess that this uncertainty will be at the few percent level.

## 11 Strategy for Background Determination

The backgrounds to this analysis are from $WW$, $WZ$, $ZZ$, $W$+jets, Drell-Yan, and QCD. As discussed in Section 2, the $N_{jets} = 0$ and 1 bins will be used to validate our background predictions.

The $WW$, $WZ$, and $ZZ$ backgrounds will come almost entirely from Monte Carlo. The Monte Carlo acceptances for these processes will be corrected for differences in data and Monte Carlo lepton identification and trigger efficiencies, as determined from the tag-and-probe method. Further corrections may have to be applied if we find additional important discrepancies, *e.g.*, in $\not\!\!E_T$. The theoretical uncertainties in the cross-section calculations will be reflected in uncertainties in the background normalization.

Another set of uncertainties will arise from the modeling of the jet multiplicity in these events, as well as the uncertainty in the jet energy scale. For the moderate jet transverse momenta and the not too high jet multiplicities relevant to our analysis, we have seen that the Pythia and Alpgen model do not differ very much, at least for Drell-Yan, see Figure 15[20].

Although the relative uncertainties on these backgrounds will be quite large, their impact on the uncertainty on the cross-section estimate will not be large. This is because these backgrounds are quite small, see Figure 11.

The Drell-Yan background is very important in the $ee$ and $\mu\mu$ case. The level of background depends on the $\not\!\!E_T$resolution in CMS. We will use the $\not\!\!E_T$distribution on $Z$ events to verify and calibrate the $\not\!\!E_T$resolution for Drell-Yan events. If we find good agreement between data and Monte Carlo for the $\not\!\!E_T$in $Z$ events, we will rely on Monte Carlo to model the $\not\!\!E_T$for events away from the $Z$ peak. On the other hand, if the $\not\!\!E_T$in $Z$ data and Monte Carlo does not agree, we will have to tune the Monte Carlo or develop ad-hoc methods to use the $\not\!\!E_T$in $Z$ data to predict the Drell-Yan background.

Just as in the case of dibosons, there will be uncertainties in the Drell-Yan background estimate due to jet multiplicity and and jet energy scale modeling. However, we will use the $Z$ peak to calibrate these properties for Drell-Yan events. The observed $Z$ cross-section will also be used to normalize the total Monte Carlo expectation.

In the case of $W$+jets and QCD, the background predictions will be extracted in a data-driven method. This method is described in Section 12

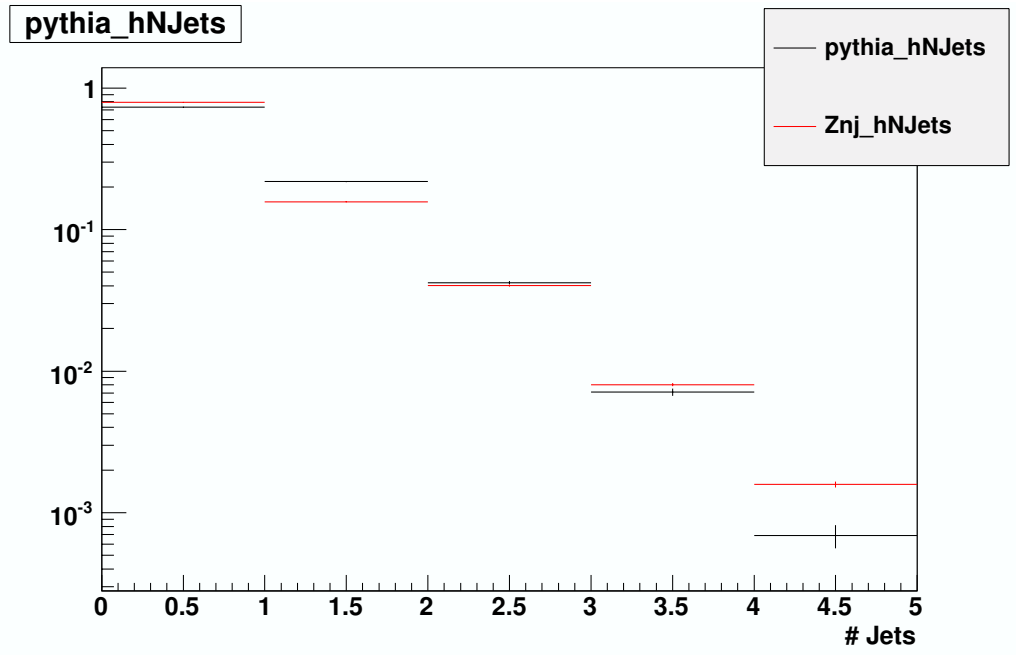



Figure 15: Jet multiplicity in $Z$ events from Pythia and Alpgen. Jets have uncorrected $P_T > 15$ GeV and $|\eta| < 3$. This study was performed in CMSSW_1_3. What is plotted is the fraction of events in each jet multiplicity bin.

# 12 Data Driven Method for Fake Lepton backgrounds

In this Section we will describe a data driven method to extract the fake lepton background contribution. The method described here has been used by CDF in a number of dilepton analyses, *e.g.*, the search for $WW$[21].

The method consists in measuring a "fake rate" in QCD events, *i.e.*, events triggered by jet triggers. The fake rate is defined as the ratio of good leptons, *i.e.*, leptons passing all the requirements, and "fakeable objects". A fakeable object is defined as a lepton candidate passing much looser requirements than the good lepton defined above. The fake rate is parametrized as a function of the relevant variables, *e.g.*, $P_T$ and $\eta$.

To estimate the fake background, the dilepton analysis is carried out by selecting one good lepton and one fakeable object. The background is then estimated by scaling each events in the lepton + fakeable object sample by the fake rate for the given fakeable object. Various corrections need to be applied to this procedure to account for higher order effects, *e.g.*, the contamination of dilepton signals to the lepton + fakeable object sample.

The key assumption is that the fake rate measured on QCD events can be applied to jets in $W+$ jets, since the major fake background comes from this sample. This is an assumption that cannot be proven rigorously.

To get a feeling for the validity of this assumption, one can apply this method to Monte Carlo $W+$ jets events. Another way to test this assumption is to compare the fake rates measured on different QCD samples, and in the photon + jet sample. All of these cross-checks can be used to assign a systematic uncertainty to this procedure.

In the remainder of this Section we will try to apply this procedure to the analysis described in this note.

**This work is in progress.**

# 13 Cross-section determination

We can now estimate the uncertainty on the cross-section determination from this analysis. Concentrating on the events with $N_{jets} \geq 2$, we project a statistical uncertainty of 9% from Figure 11 (this is for all three modes combined). Systematic uncertainties will arise from the imperfect knowledge of the lepton reconstruction efficiencies, the jet energy scale, the background subtraction, the Monte Carlo modeling of $t\bar{t}$ production, and the luminosity normalization.

Efficiency of lepton selection (by trigger and reconstruction) will be measured using the tag-and-probe method from the large sample of $Z$ decays. We expect systematic uncertainty on this measurement to be at the few percent level.

Uncertainty in the jet energy scale of 10%, as discussed in Section 9 translates into acceptance uncertainty of 5%. Note that in the same section we have also seen that there is a $\approx 5\%$ loss of efficiency per electron in using 10 $\mathrm{pb}^{-1}$ vs. 100 $\mathrm{pb}^{-1}$ calibrations. We do not regard this as a systematic uncertainty: whatever the electron efficiency is, it will be **measured** by tag-and-probe, with its own statistical and systematic uncertainties.

Assuming the current understanding of the background, the signal-to-noise is expected to be of order 8 to 1. We conservatively guess that we will know it to 50% of itself; this translates into a systematic uncertainty of 6% on $\sigma(t\bar{t})$. Alternatively, we could think of measuring $\sigma(t\bar{t})$ using $e\mu$ events only. Then, the statistical uncertainty would be 13%, the signal-to-noise would be of order 27 to 1, and the systematic uncertainty on the background subtraction would be negligible on this scale.

There will be a systematic uncertainty associated with the modeling of $t\bar{t}$ production. In the early days of CMS this will be assessed by running different physics models, *e.g.*, Alpgen vs. Pythia vs. MC@NLO. We do not currently have an estimate of this uncertainty, but we do not expect it to be large.

Finally, there will be an uncertainty associated with the luminosity normalization. This may very well turn out to be the largest one.

Putting all of these together, we estimate a statistical uncertainty in the $t\bar{t}$ cross-section of 9% and a systematic uncertainty of order 9%. This last uncertainty does not include the luminosity uncertainty

# 14 Conclusions

We have presented expectations of observing top quark pair production in a final state with two leptons with high transverse momentum and jets using the first $10\ \mathrm{pb}^{-1}$ of CMS data. The event selection algorithm we employ is found to be robust against varying detector performance which may be expected in the early operation of CMS. Clear observation of the signal is expected in the sample with two or more jets with a signal-to-noise ratio of about 7 to 1 in all channels combined and about 25 to 1 in the $e\mu$ channel alone. With the first $10\ \mathrm{pb}^{-1}$ of data we expect to measure the top pair production cross section with the statistical uncertainty of $9\%$ using all dilepton channels or with an uncertainty of $13\%$ using $e\mu$ channel alone. Based on the limited study of systematic effects we expect the systematic uncertainty of order $9\%$ (excluding uncertainty on integrated luminosity of the sample).

In the sample of CMS data we anticipate to use for analysis we expect to have controls over effects not necessarily predicted well by the simulation. Events with 0 and 1 jets not dominated by contributions from top-quark pair production will be used to check our expectations of background events. The lepton selection efficiencies will be measured from $Z \to \ell\ell$ events in the data. We are also developing methods to estimate backgrounds arising from the $W$+jets or QCD multijet processes.

# References


[1] Link to the note by the Oviedo and Strasbourg groups.

[2] Link to some web page (twiki) that summarizes the CSA07 data sets.

[3] `https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopQuarkCSA07`.

[4] `https://twiki.cern.ch/twiki/bin/view/CMS/CSA07GeneratorInformation`.

[5] The CSA07 spreadsheet prepared by S. Lowette,
`UserCode/lowette/CSA07EffAnalyser/CSA07EffAnalyser/test/csa07skimeffanalysis.xls`,

gives the cross-sections for $t\bar{t}$, $W$, and Drell-Yan as 837 pb, 58247 pb, and 5771 pb, respectively. However, the assumed cross-section in the `CSA07EventWeightProducer` module for $t\bar{t}$ is 447 pb. The 837 pb is the NLO cross-section, while the 447 pb is the LO cross-section. This has been discussed in `https://hypernews.cern.ch/HyperNews/CMS/get/generators/202/1/1.html`, for example.

[6] D. Acosta *et al.*, CMS AN 2007/009.

[7] Presentation by J. Cuevas at the 4 March 2008 Top PAG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=29778`, and Analysis Note in preparation.

[8] Presentation by V. Krutelyov at the 25 January 2008 Top PAG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=27655`.

[9] Presentation by C. Campagnari at the 18 September 2007 Muon POG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=20377`.

[10] Presentation by M. Sani at the 24 October 2007 Egamma POG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=22460`.

[11] Presentation by C. Campagnari at the 19 February 2008 Top PAG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=27807`.

[12] Presentation by P. Kalavase at the 14 January 2008 Egamma POG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=26395`.

[13] Presentation by P. Kalavase at the 26 February 2008 Egamma POG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=29233`.

[14] Presentation by C. Campagnari at the 19 February 2008 Top PAG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=27807`.

[15] D. Elvira, `https://hypernews.cern.ch/HyperNews/CMS/get/met/12.html`.

[16] G. Landsberg, `https://hypernews.cern.ch/HyperNews/CMS/get/ewk-muons/46/2/1/1.html`.

[17] Find some reasonable reference to tag-and-probe for both electrons and muons.

[18] `ppMuPt20-15/CMSSW_1_6_7-CSA07-1205151815/AODSIM`.

[19] Presentation by J. Lamb at the 11 February 2008 egamma POG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=27542`.

[20] Presentation by J. Ribnik at the 9 October 2007 EWK PAG meeting, `http://indico.cern.ch/conferenceDisplay.py?confId=22820`.

[21] The CDF dilepton WW analysis paper.